Reservoir Computing Prediction of the Spread of COVID-19

Jacob Jeffries[1], Declan Norton[2], Tom Antonsen[2], Michelle Girvan[2], Brian Hunt[2], Andrew Pomerance[3]

[1]Clemson University [2]University of Maryland, College Park [3]Potomac Research

Summary

- Accurate COVID-19 forecasting leads to better pandemic-related decision making
- Reservoir computing is an efficient and powerful tool for time series prediction
- By learning the history of the pandemic, a reservoir computer (RC) can accurately forecast incident COVID-19 cases **weeks** in advance
- An RC only needs to have knowledge about the near-peak behavior of incident cases to make accurate predictions

Introduction

- **Accurate** and **computationally efficient** forecasts of COVID-19 are imperative to making quick, well-informed decisions about the pandemic for policymakers, healthcare infrastructure, and even as an individual.
- Accurate forecasting is difficult due to the complex, non-stationary nature of disease dynamics (emerging variants, interventions, and more), as well as noisy, inconsistent reporting (Figure 1).

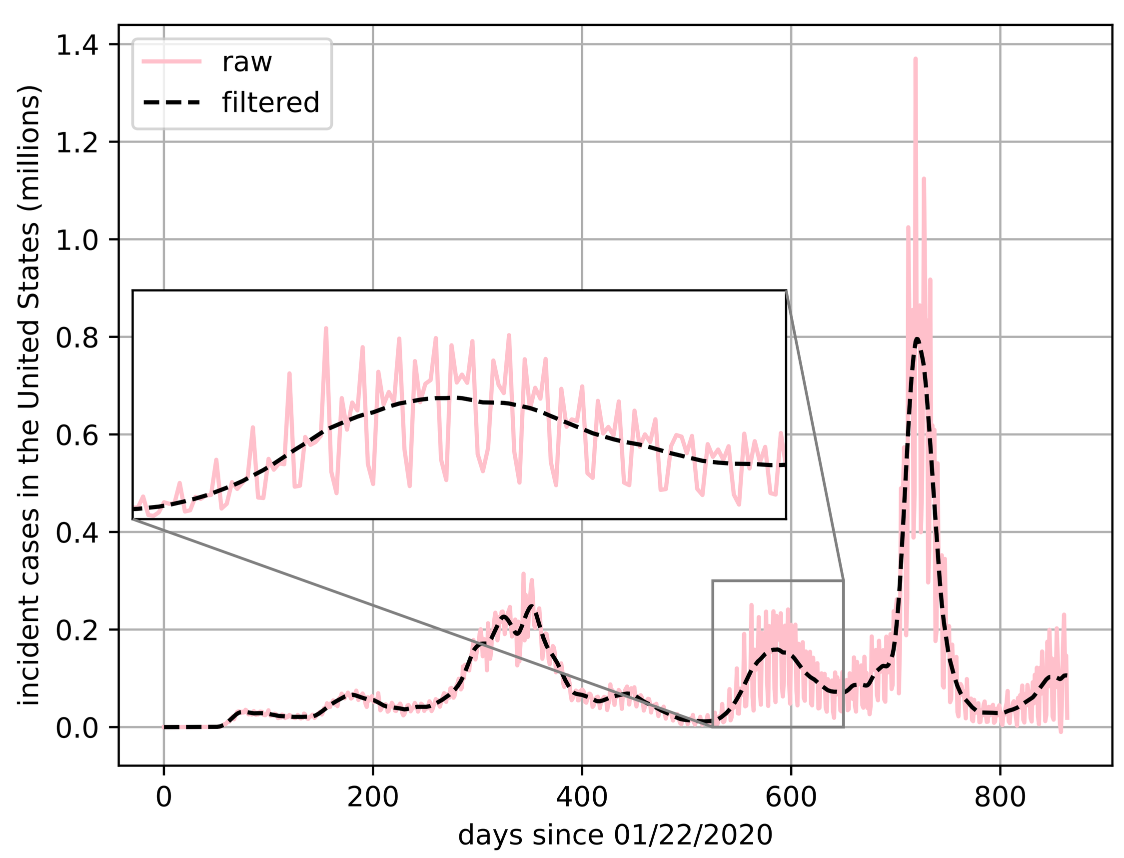

Figure 1: Incident COVID cases in the United States since January 22, 2020 [2]. Filter applied to account for inconsistent, error-prone reporting.

- Reservoir computing is an efficient and accurate machine learning framework for efficient time series prediction.

Reservoir Computing

- An RC maps an input signal (u_t) to a reservoir state (r_t), and then an output layer is trained to recover u_t from r_t over data within some training period (Figure 2).

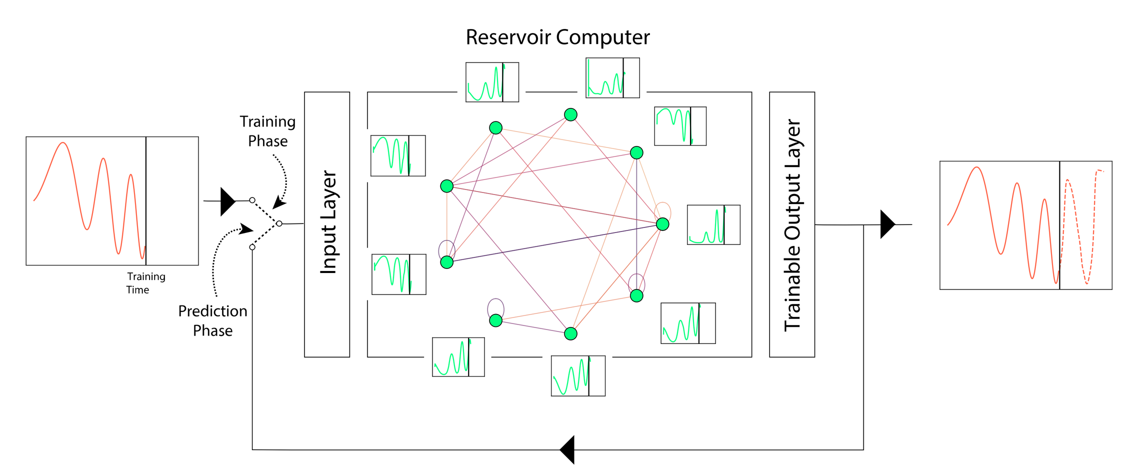

Figure 2: Traditional reservoir computing schematic

- For $t \leq T$ (training time), the output layer is trained such that the RC recovers the input.
- For $t > T$, the reservoir state autonomously evolves, and is then fed into the output layer, predicting u_t.
- The output layer is chosen to be linear ($u_t = W r_t$)

Fitting to the SIR model

In the susceptible-infected-recovered (SIR) model:

$$S_{t+1} = S_t - \beta \Delta t S_t I_t$$
$$I_{t+1} = I_t + \beta \Delta t S_t I_t - \gamma \Delta t I_t$$
$$R_{t+1} = R_t + \gamma \Delta t I_t$$

the incident case time series is $S_t - S_{t+1}$, where β and γ are the contact and recovery rate respectively. We fit the SIR model over a wave near each peak in the smoothed incident case time series (x_t) identified via a wavelet transform [3]. We also fit for the initial conditions S_0 and I_0 (Figure 3).

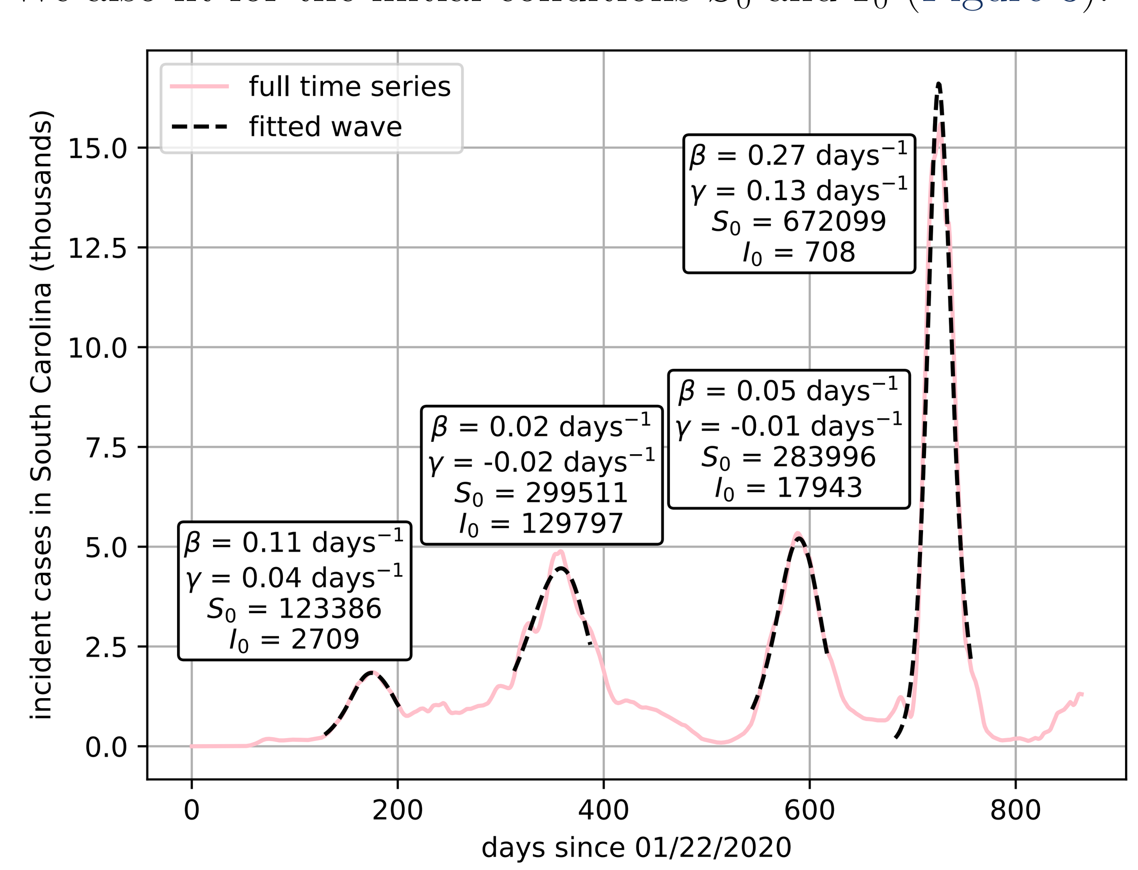

Figure 3: (x_t) for South Carolina in pink, fit for constituent waves shown in black.

Preprocessing and Training

Training signals were either generated by waves:

- Picking a training time
- Fitting all pre-training time waves (indexed by k) to the SIR model
- Designating all signals with accurate SIR fits (max percent error $< 50\%$) to be training waves
- Each detrended training wave is then:

$$y_t^{(k)} = \log\left(x_{t+1}^{(k)} / x_t^{(k)}\right)$$

Or by all available data:

- Pick a training time
- Segment all pre-training time signals into 75 day long windows
- Designating all detrended segments to be training waves

Multiple RCs were then trained on all detrended training waves in both cases. We chose an RC architecture that has shown predictive strength for chaotic systems [4].

Results

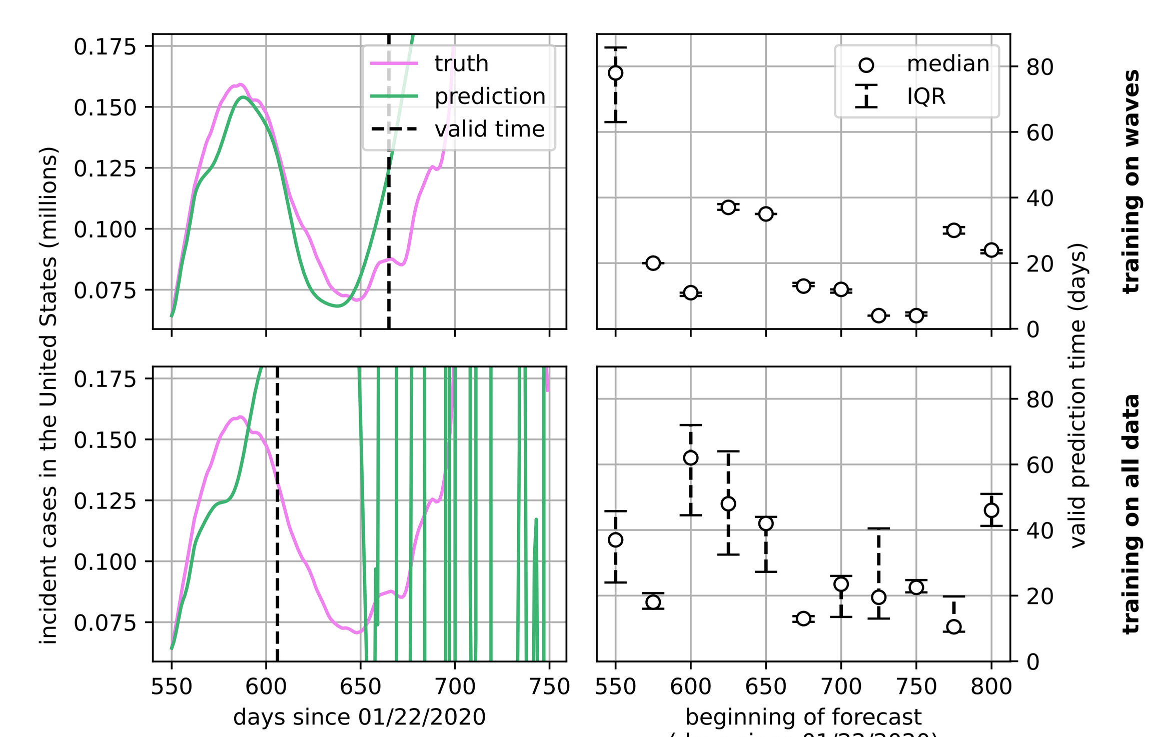

Figure 4: Example predictions in the United States (left) beginning the forecast at 07/25/2021, and valid prediction times at multiple training times (right) by training on windows (top) and training on all data (bottom). Inaccuracy tolerance of 40%.

- An RC can predict incident COVID-19 cases **weeks into the future**
- Similar predictive power when training on waves vs. training on all data (Figure 4)
- Very poor predictive power (on the order of 4-5 days in the very best case scenarios) when only training and predicting on singular states (not shown).

Conclusion

- Reservoir computing accurately forecasts COVID-19 **weeks into the future** using the history of the pandemic in all states and a two-week resync period
- Training on wave-like portions of the pandemic yields similar predictive strength compared to training on all available data
- This beats a large complication of other models, which only confidently reports one-week forecasts [2]

Future Work

- Couple the SIR model with the hidden state of an RC [4]
- Build a meta-learning approach [1] to predict at the state level by learning the properties of RCs trained on individual windows

References

[1] Daniel Canaday, Andrew Pomerance, et al. A meta-learning approach to reservoir computing: Time series prediction with limited data. *arXiv*, 2021.

[2] Estee Y Cramer, Yuxin Huang, et al. The united states covid-19 forecast hub dataset. *medRxiv*, 2021.

[3] Pan Du, Warren A. Kibbe, et al. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 2006.

[4] Jaideep Pathak, Alexander Wikner, et al. Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos*, 2018.

Acknowledgements

This work was sponsored by the National Science Foundation Award Number PHY2150399

Contact Information

- Email: jwjeffr@clemson.edu

CLEMSON

TREND REU

Training and Research Experiences in Nonlinear Dynamics